

- 1. The big deal about big data**
- 2. From Operational Data to Trusted Big Data**

IBM Information Governance Platform

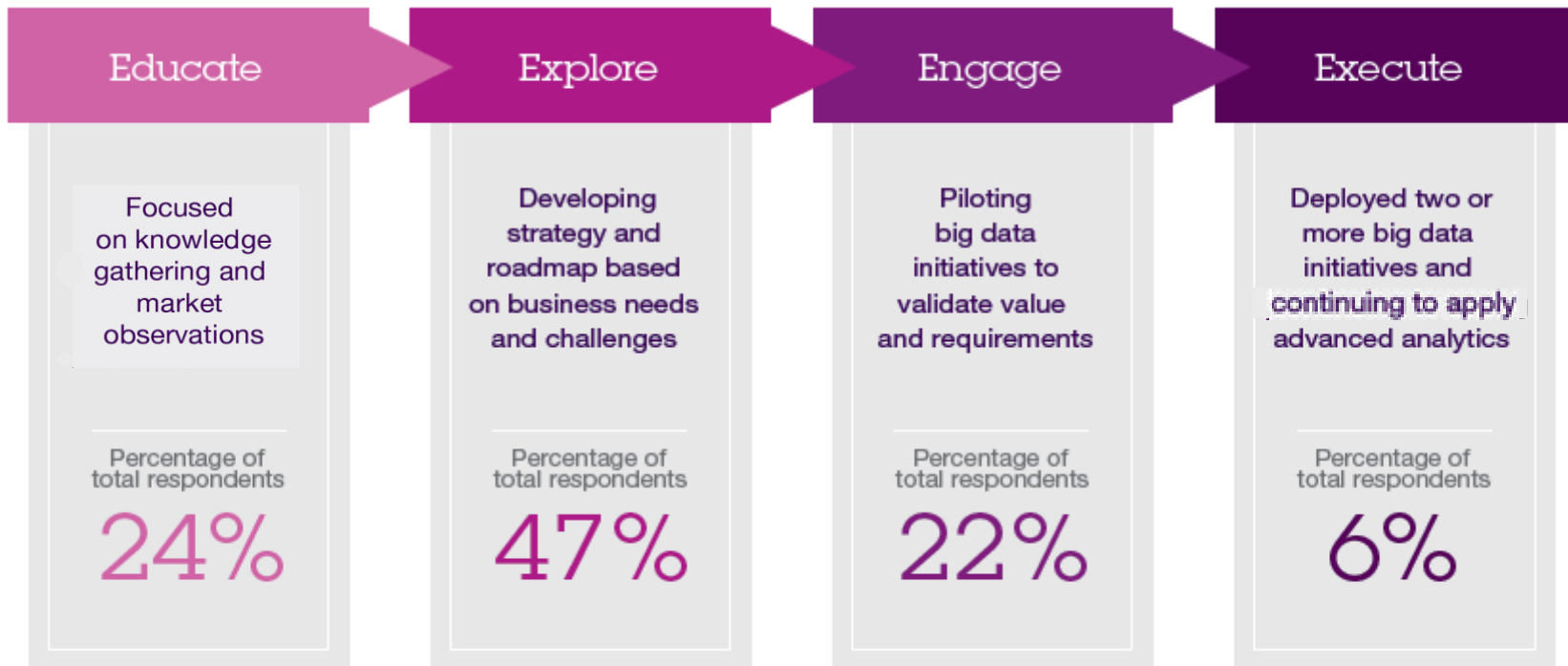
Andre De Locht
Information Integration
Sr Business Consultant
☎ +32 476 870 354
✉ andre.de.locht@be.ibm.com

IBM delivers a governable,
consumable Big Data platform
that's steeped in analytics for data in-
motion and data at-rest.

A recent Institute for Business Value study highlights how organizations are adopting big data in four phases



Big data adoption



When segmented into four groups based on current levels of big data activity, respondents showed significant consistency in organizational behaviors

Total respondents n = 1061

Totals do not equal 100% due to rounding



C&A Brasil

about an hour ago via PostCron

Calças com a barra dobrada continuam sendo um sucesso entre os homens. Abuse! — with Michel Almeida.



Like · Comment · Share

35

Giselle Magallanes, Cesar Santos and 213 others like this.

View all 27 comments



Fuchs Erick e pra os que ficam falando de 'isso eh a ultima moda na europa", homophobic seal pra vcs ahuaahuahhuahua

31 minutes ago

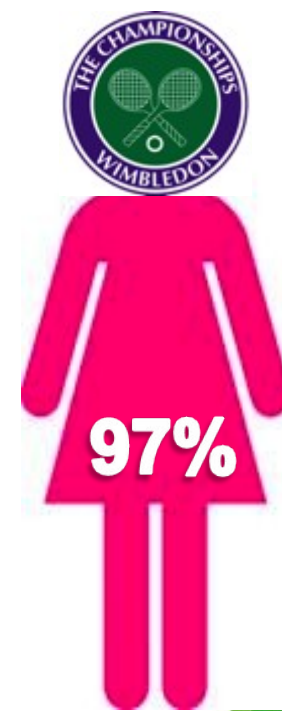
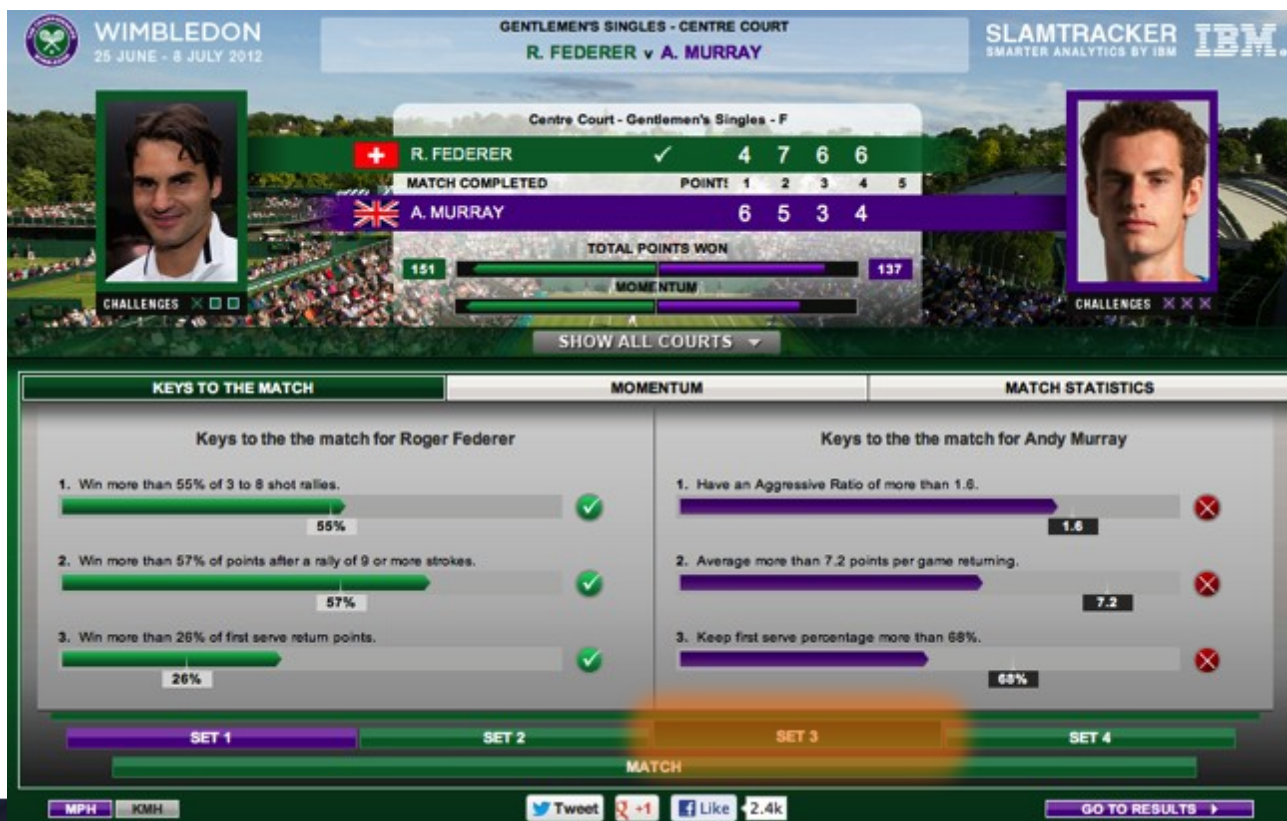


Giovane Farias a inveja tem facebook galera!

20 minutes ago

SlamTracker at Wimbledon

- Combines a corpus of data points made up of a player's historical winning performance, correlated with winning statistics of the last 5 grand slams, to create "Keys to the Match" win prediction engine



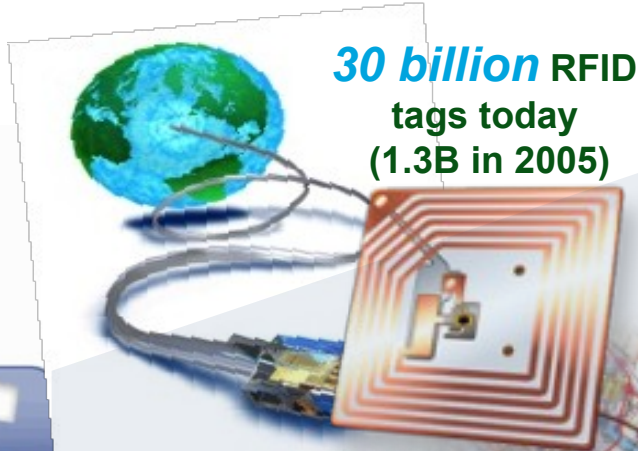
The Social Layer in an Instrumented Interconnected World

12+ TBs



500+ TBs

? TBs of data every day



30 billion RFID tags today (1.3B in 2005)



4.6 billion camera phones world wide



100s of millions of GPS enabled devices sold annually



76 million smart meters in 2009... 200M by 2014



2+ billion people on the Web by end 2011



http

The Big Data Conundrum



- **The economies of deletion have changed....**
 - Leading us into new opportunities and challenges
- **The percentage of available data an enterprise can analyze is decreasing proportionately to the amount of data available to that enterprise**
 - Quite simply, this means as enterprises, we are getting “more naive” about our business over time
- **Just collecting and storing “Big Data” doesn’t drive a cent of value to an organization’s bottom line**

Data AVAILABLE to
an organization

Signals
and
Noise



Data an organization
can PROCESS

Applications for Big Data Analytics

Smarter Healthcare



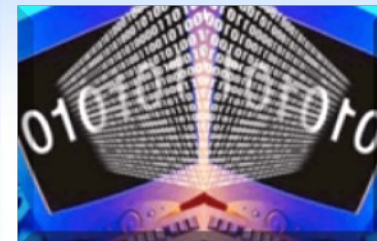
Multi-Channel



Finance



Log Analysis



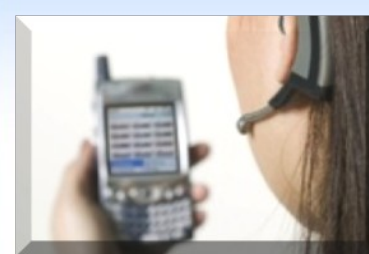
Homeland Security



Traffic Control



Telecom



Search Quality



Manufacturing



Trading Analytics



Fraud and Risk



Retail: Churn, NBO



5 Big Data Patterns



Big Data Exploration
Find, visualize, understand all big data to improve business knowledge



Enhanced 360° View of the Customer
Achieve a true unified view, incorporating internal and external sources



Security/Intelligence Extension
Lower risk, detect fraud and monitor cyber security in real-time



Operations Analysis
Analyze a variety of machine data for improved business results



Data Warehouse Augmentation
Integrate big data and data warehouse capabilities to increase operational efficiency



Big Data Exploration

Find, visualize, understand
all big data to improve
business knowledge



- **Exploring 4 TB to drive point business solutions (supplier portal, call center, etc.)**
- **Single-point of data fusion for all employees to use**
- **Reduced costs & improved operational performance for the business**



Enhanced 360° View of the Customer

Achieve a true unified view,
incorporating internal and
external sources

*Confidential,
Internal Use Only*

**US Insurance Company
Anonymized on client request**

- **Increase revenue and decrease cost in the call center**
- **Increase customer & employee satisfaction**
- **Leverage new data types in customer analysis**



Security/Intelligence Extension

Lower risk, detect fraud and monitor cyber security in real-time



Asian Government Agency

Blinded for confidentiality

National Intelligence Platform Extension

Using InfoSphere Streams to filter & analyze all Internet traffic (social media, email, etc) to track persons of interest (drug/sex traffickers, terrorists, illegal refugees/immigrants) and civil/border activity.



- Intelligent Infrastructure Management: log analytics, energy bill forecasting, energy consumption optimization, anomalous energy usage detection, presence-aware energy management
- Optimized building energy consumption with centralized monitoring; Automated preventive and corrective maintenance
- Utilized InfoSphere Streams, InfoSphere BigInsights, IBM Cognos



Operations Analysis

Analyze a variety of machine data for improved business results

**Automotive industry
Anonymised on Client
request**

- **Creates pre-processing hub and performs ad hoc analysis**
- **Hadoop-based landing zone used to store, manage and analyze structured, semi-structured and multi-structured data before moving to the warehouse**
- **Benefits: Data warehouse optimized for workload and performance**
- **Utilized InfoSphere BigInsights, InfoSphere DataStage**



Data Warehouse Augmentation
Integrate big data and data warehouse capabilities to increase operational efficiency

In Order to Realize New Opportunities, You Need to Think Beyond Traditional Sources of Data

Transactional and Application Data



- Volume
- Structured
- Throughput

Machine Data



- Velocity
- Semi-structured
- Ingestion

Social Data



- Variety
- Highly unstructured
- Veracity

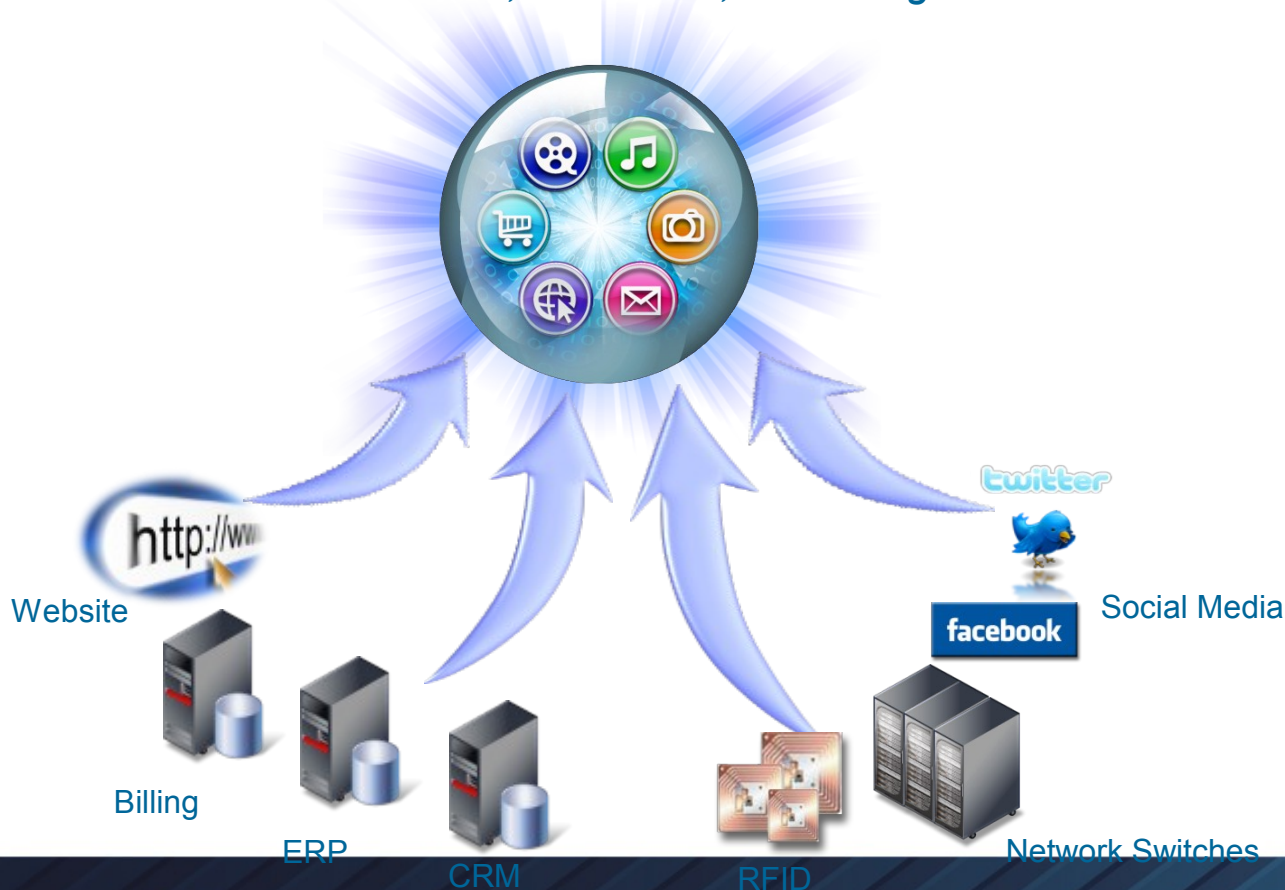
Enterprise Content



- Variety
- Highly unstructured
- Volume

Big Data Technology Makes it Possible to Analyze ALL Available Data

Cost effectively manage and analyze
all available data in its native form
unstructured, structured, streaming



The Big Data Platform Manifesto

**Understand and Navigate
Federated Big Data Sources**



**Federated Discovery
and Navigation**

**Manage and Store Huge
Volume of any Data**



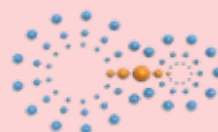
**Hadoop File System
MapReduce**

Structure and Control Data



Data Warehousing

Manage Streaming Data



Stream Computing

Analyze Unstructured Data



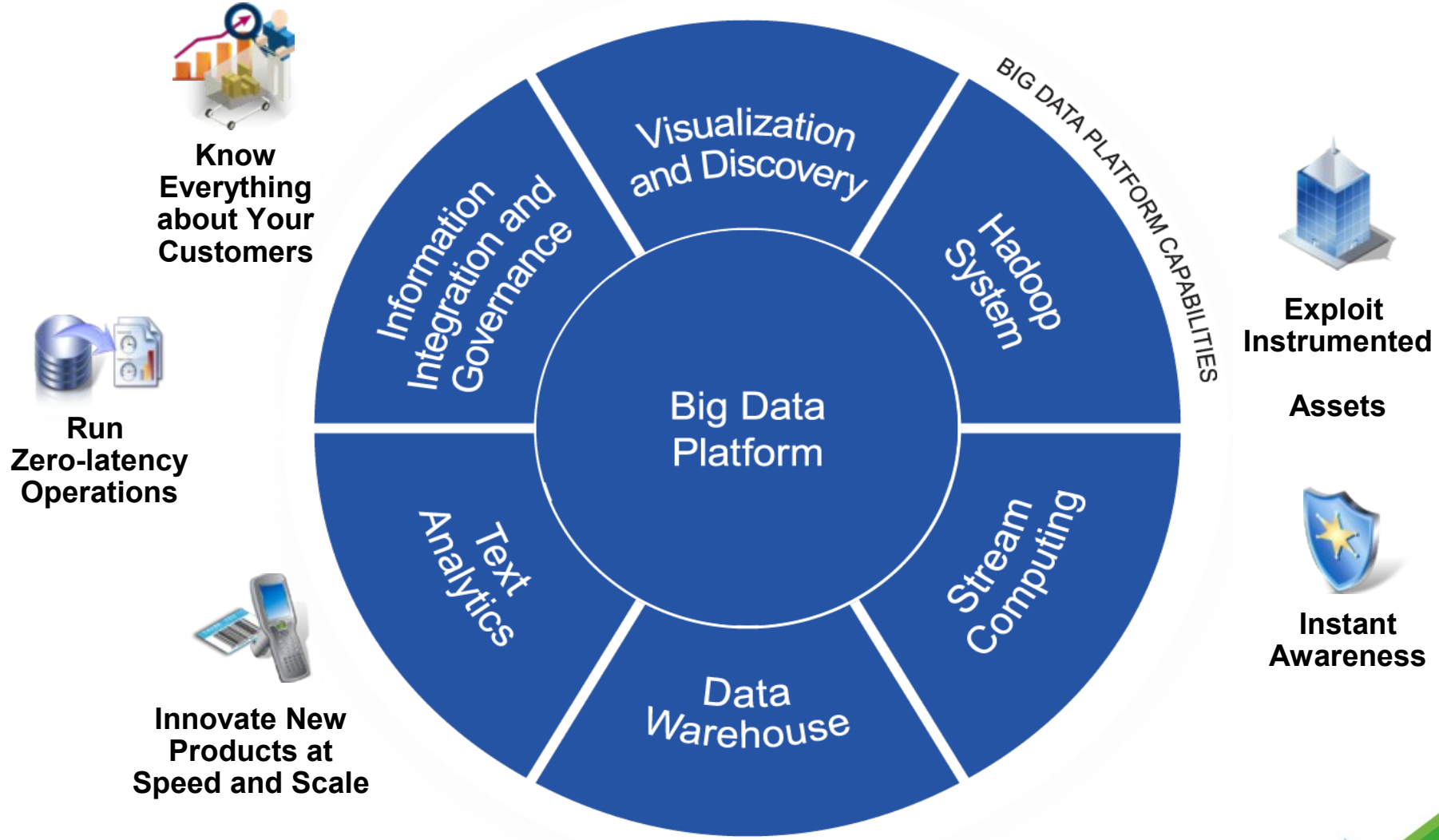
Text Analytics Engine

**Integrate and Govern
all Data Sources**



**Integration, Data Quality,
Security, ILM, MDM**

Entry Points to a Big Data Project



Big Data Platform

Move the Analytics Closer to the Data

New analytic applications drive the requirements for a big data platform

- Integrate and manage the full variety, velocity and volume of data
- Apply advanced analytics to information in its native form
- Visualize all available data for ad-hoc analysis
- Development environment for building new analytic applications
- Workload optimization and scheduling
- Security and Governance



Big Analytics



IBM Big Data Platform

Visualization & Discovery

Application Development

Systems Management



Accelerators

Hadoop System



Stream Computing



Data Warehouse



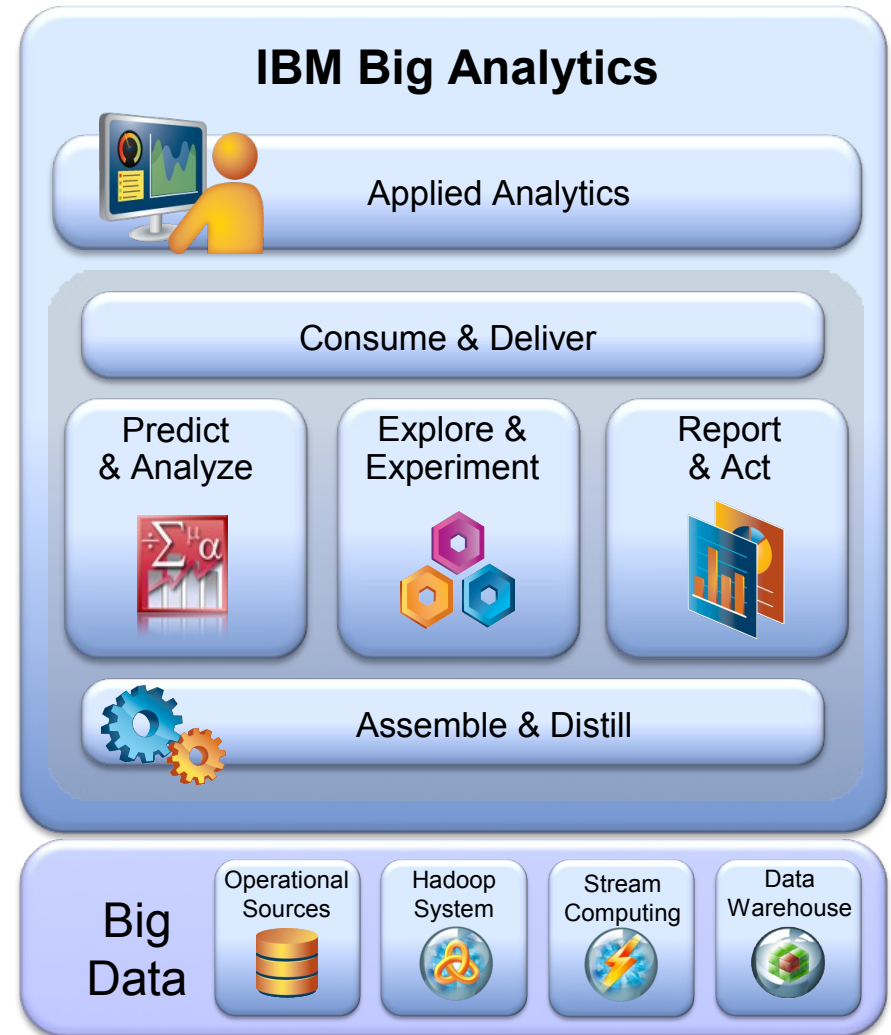
Information Integration & Governance

IBM Big Analytics



Next wave of analytics harnesses the value of the new mix of information

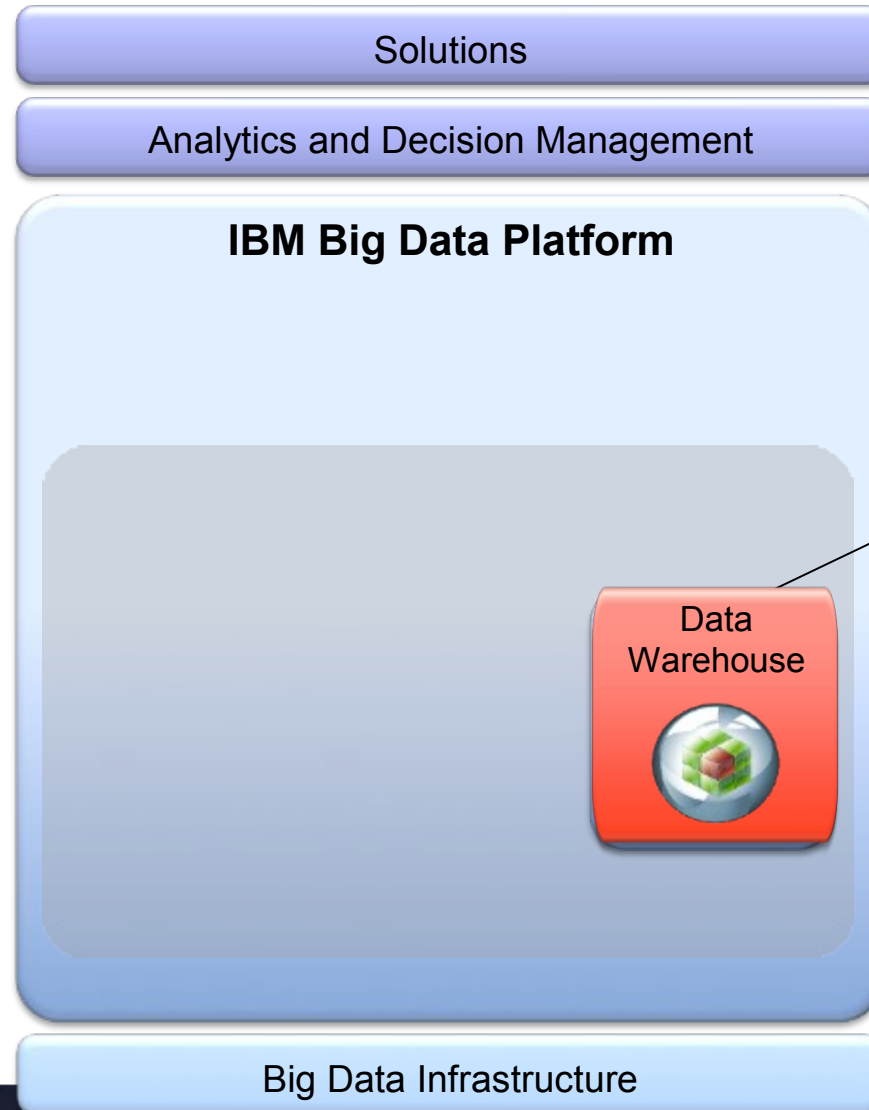
- Visualize and explore the variety, velocity and volume of big data
- Apply advanced analytics to uncover patterns previously hidden
- Blend traditional structured information with data previously unavailable
- Optimize access and delivery to take insight to action
- Extend existing capabilities to address specific analytic applications



The IBM Big Data Platform



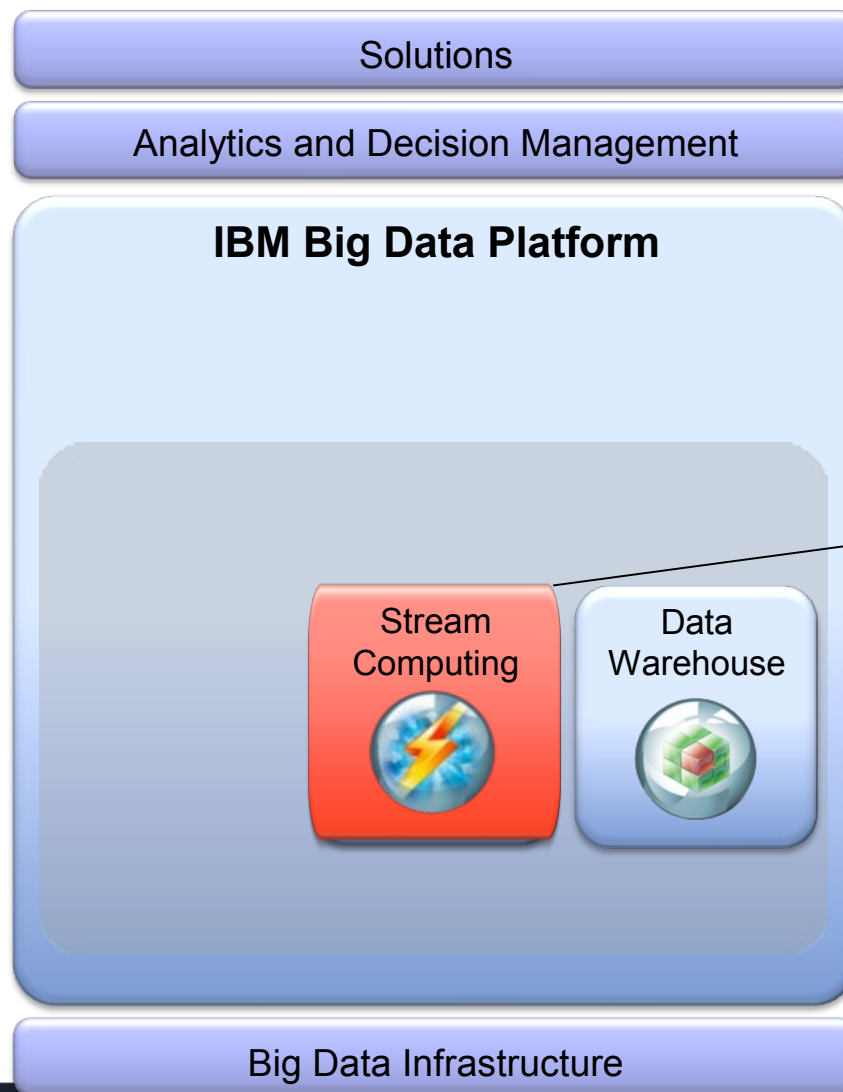
The IBM Big Data Platform



Delivers deep insight with advanced in-database analytics & operational analytics

- **PureData System** – expert integrated systems to make deep and operational analytics faster & simpler
- **InfoSphere Warehouse** -- data warehouse software to access operational info in real time

The IBM Big Data Platform



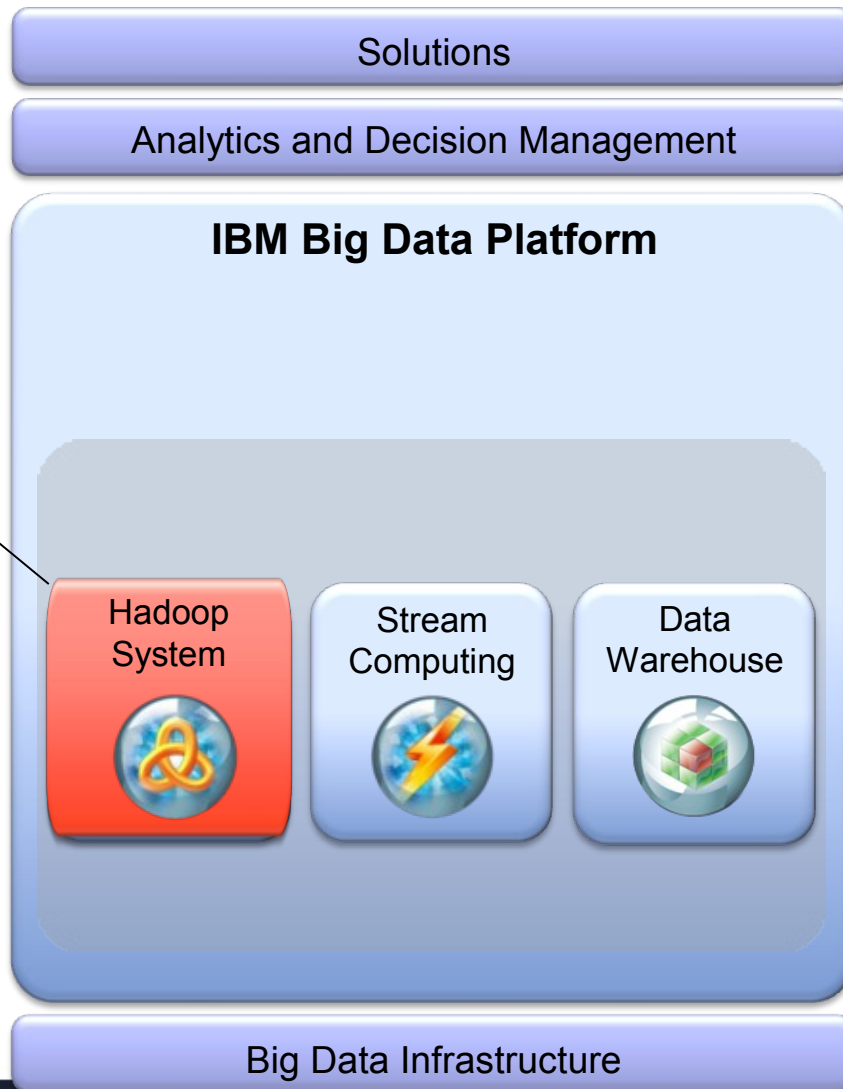
Analyze streaming data and large data bursts for real-time insights

- **InfoSphere Streams**
 - software enabling continuous analysis of massive volumes of streaming data with sub-millisecond response times

The IBM Big Data Platform

Cost-effectively analyze Petabytes of unstructured and structured data

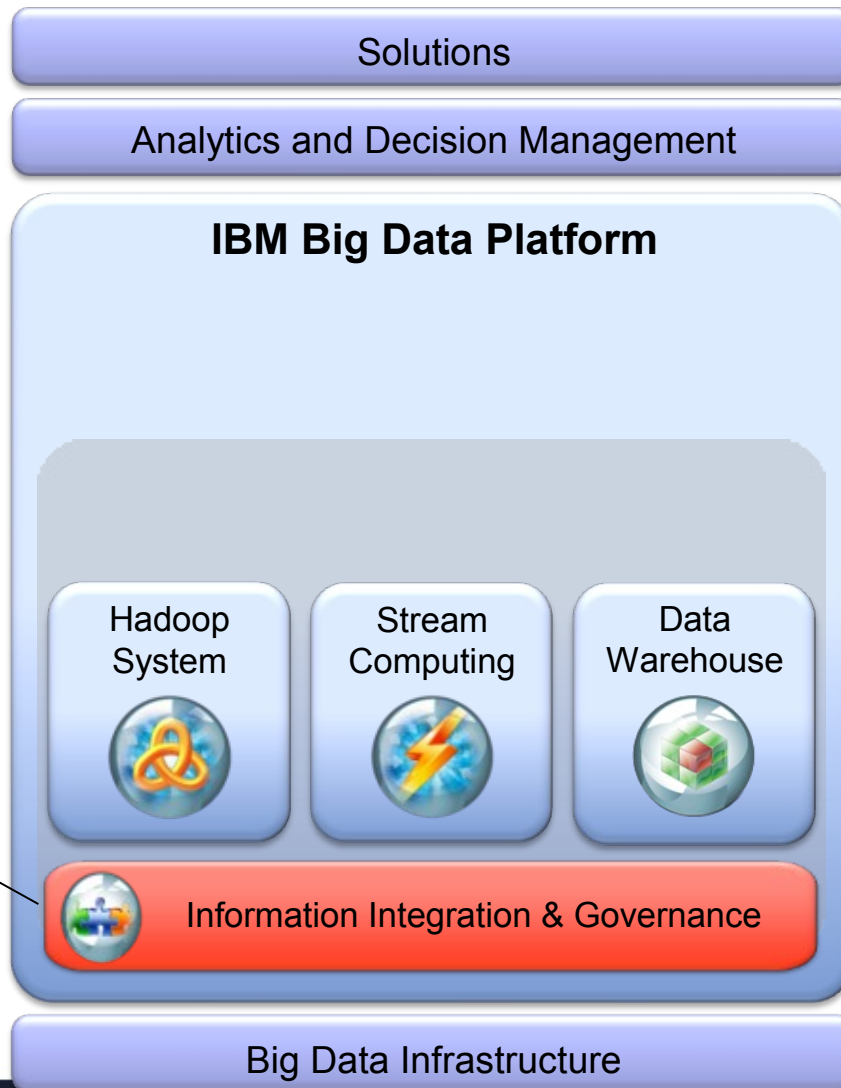
- **InfoSphere BigInsights**
--enterprise-grade Hadoop system enhanced with advanced text analytics, data visualization, tools, & performance features for analyzing massive volumes of structured and unstructured data.



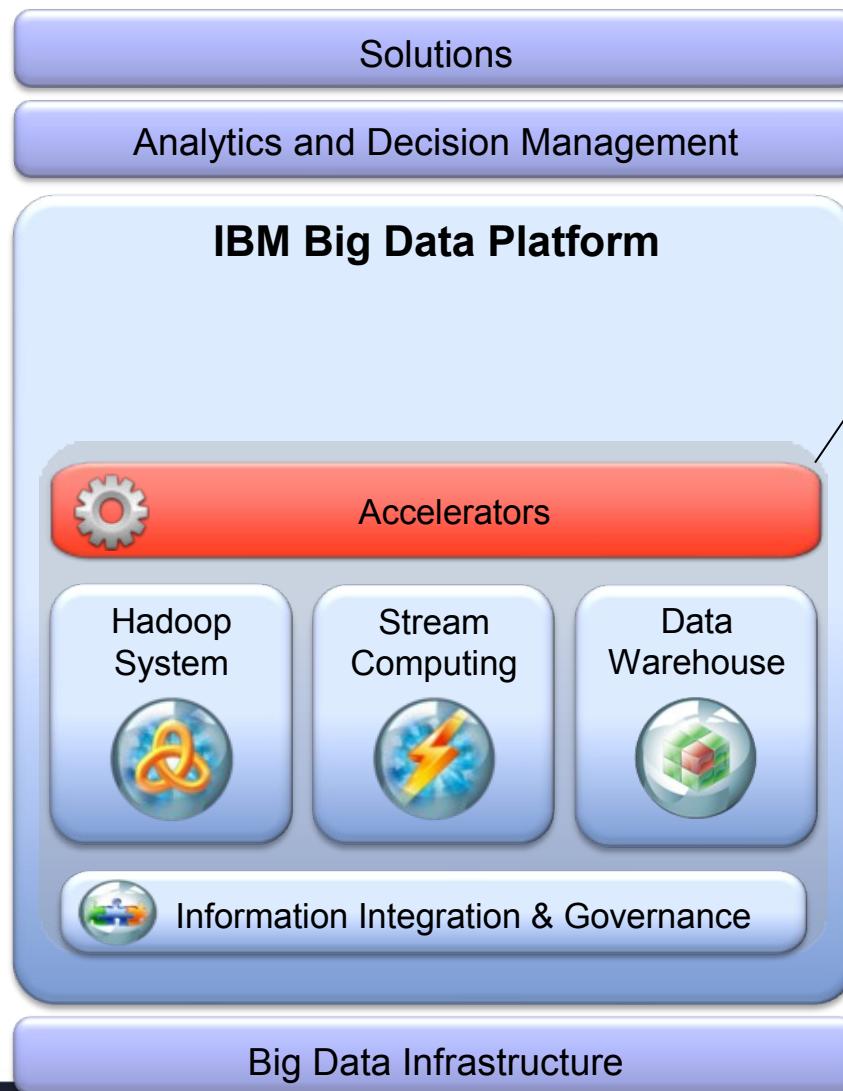
The IBM Big Data Platform

Govern data quality and manage the information lifecycle

- **InfoSphere Information Server** – Cleanses data, monitors quality and integrates big data with existing systems
- **InfoSphere Optim** – manages business information throughout its lifecycle
- **InfoSphere Master Data Management** – manages and maintains trusted views of master and reference data
- **InfoSphere Guardium** – real-time database security and monitoring



The IBM Big Data Platform



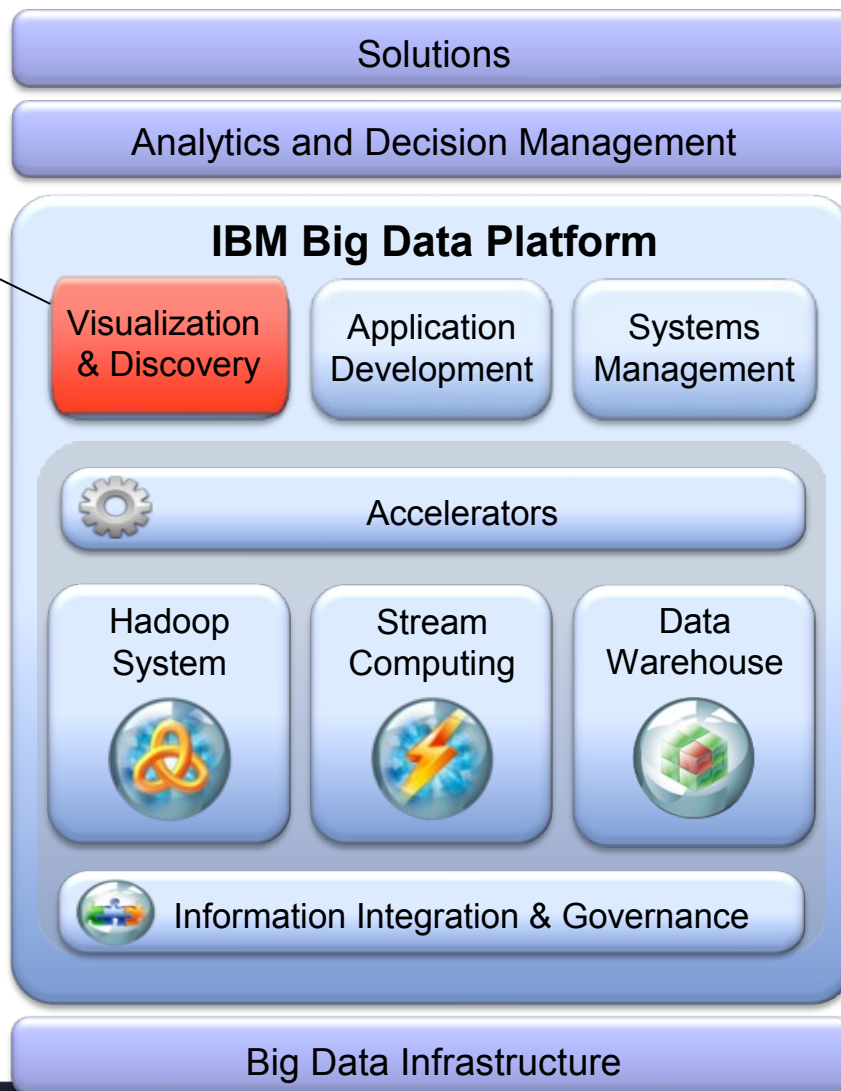
Speed time to value with analytic and application accelerators

- **Analytic Accelerators** – text analytics, geospatial, time-series, data mining
- **Application Accelerators** – financial services, machine data, social data, Telco event data
- **Industry Models** – comprehensive data models based on deep expertise and industry best practice

The IBM Big Data Platform

Discover, understand, search, and navigate federated sources of big data

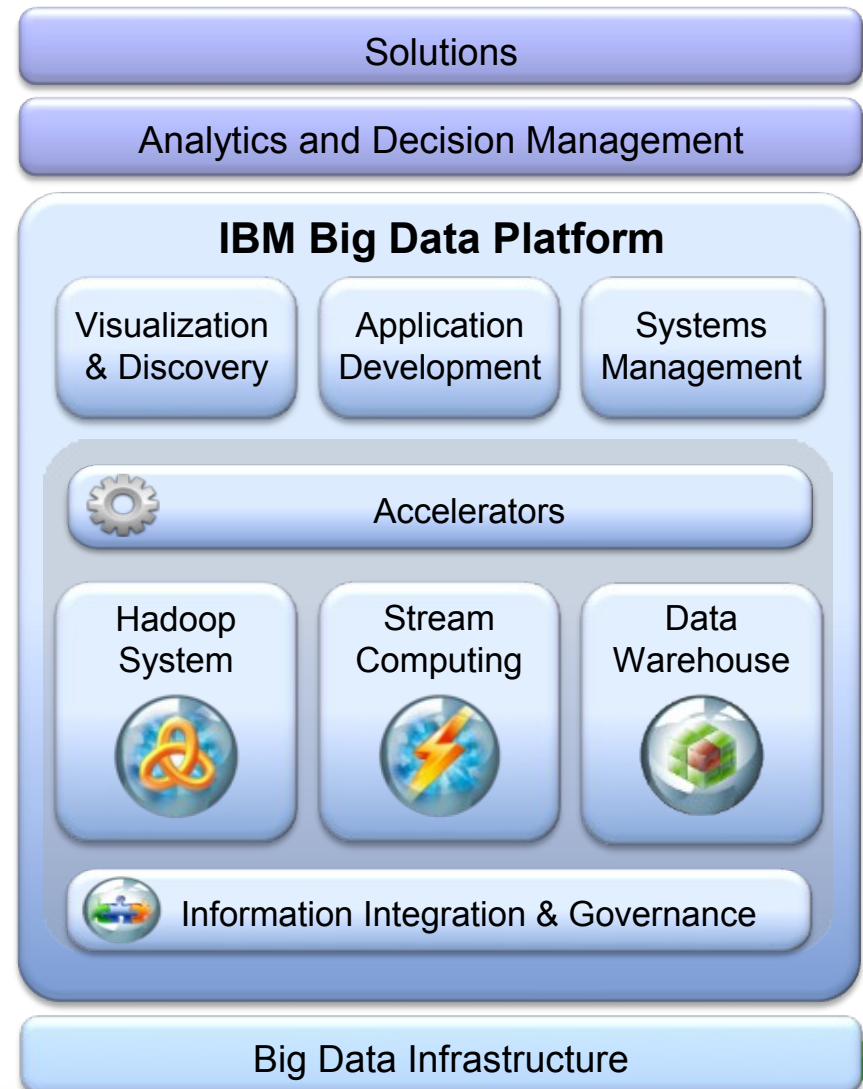
- **InfoSphere Data Explorer** – Discovery and navigation software that provides real-time access and fusion of big data with rich and varied data from enterprise applications for greater insight



The IBM Big Data Platform

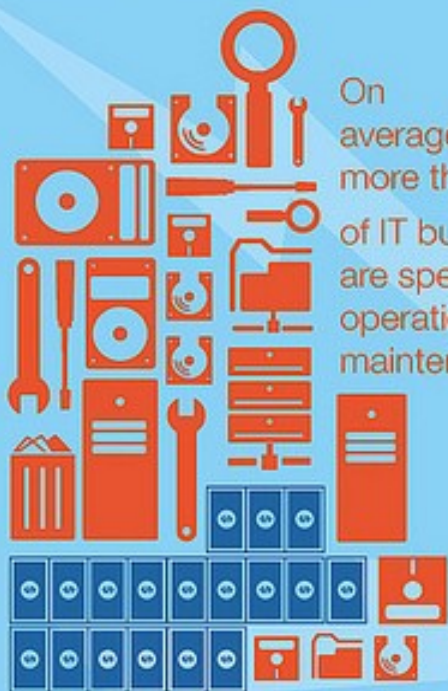


- Process **any type of data**
 - Structured, unstructured, in-motion, at-rest
- Built-for-purpose **engines**
 - Designed to handle different requirements
- Analyze **data in motion**
- **Manage and govern data** in the ecosystem
- Enterprise **data integration**
- **Grow and evolve** on current infrastructure



IT headaches

IT professionals face unnecessary delays and costs as they deploy, maintain and update their capabilities.



On average, more than **70%** of IT budgets are spent on operations and maintenance*

70%

It can take up to

4 to 6 months

just to establish hardware and software infrastructure**

Jan

Feb

Mar

Apr

May

June

July

Aug

Sept

Oct

Nov

Dec



55% of IT professionals experience downtime, that can last from anywhere between minutes to over a week when performing an infrastructure upgrade**

Nearly

2/3 of organizations fall behind schedule when deploying new IT capabilities***



Sources: * IDC, Analyst Matt Eastwood, IDC Directions Presentation, 2011

** From a commissioned study conducted by Forrester Consulting on behalf of IBM in 2011

*** IBM Market Insights Study - 2011 Business Benchmarking Time-to-Value Study

IBM

A New Family of Expert Integrated Systems

PureSystems

Systems with integrated expertise and built for cloud



Built-in Expertise

Capturing and automating what experts do – from the infrastructure patterns to the application patterns

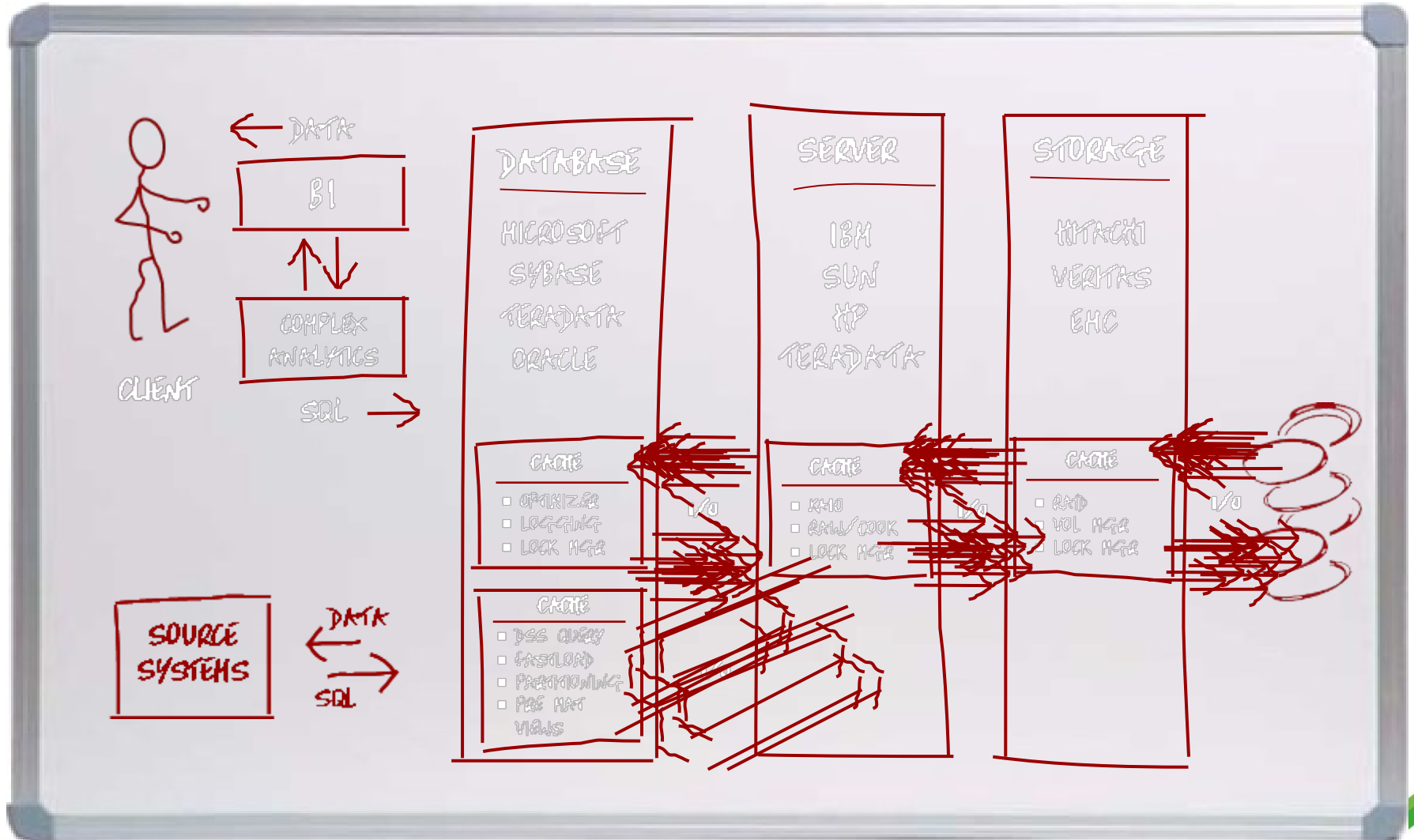
Integration by Design

Deeply integrating and tuning hardware and software – in a ready-to-go workload optimized system

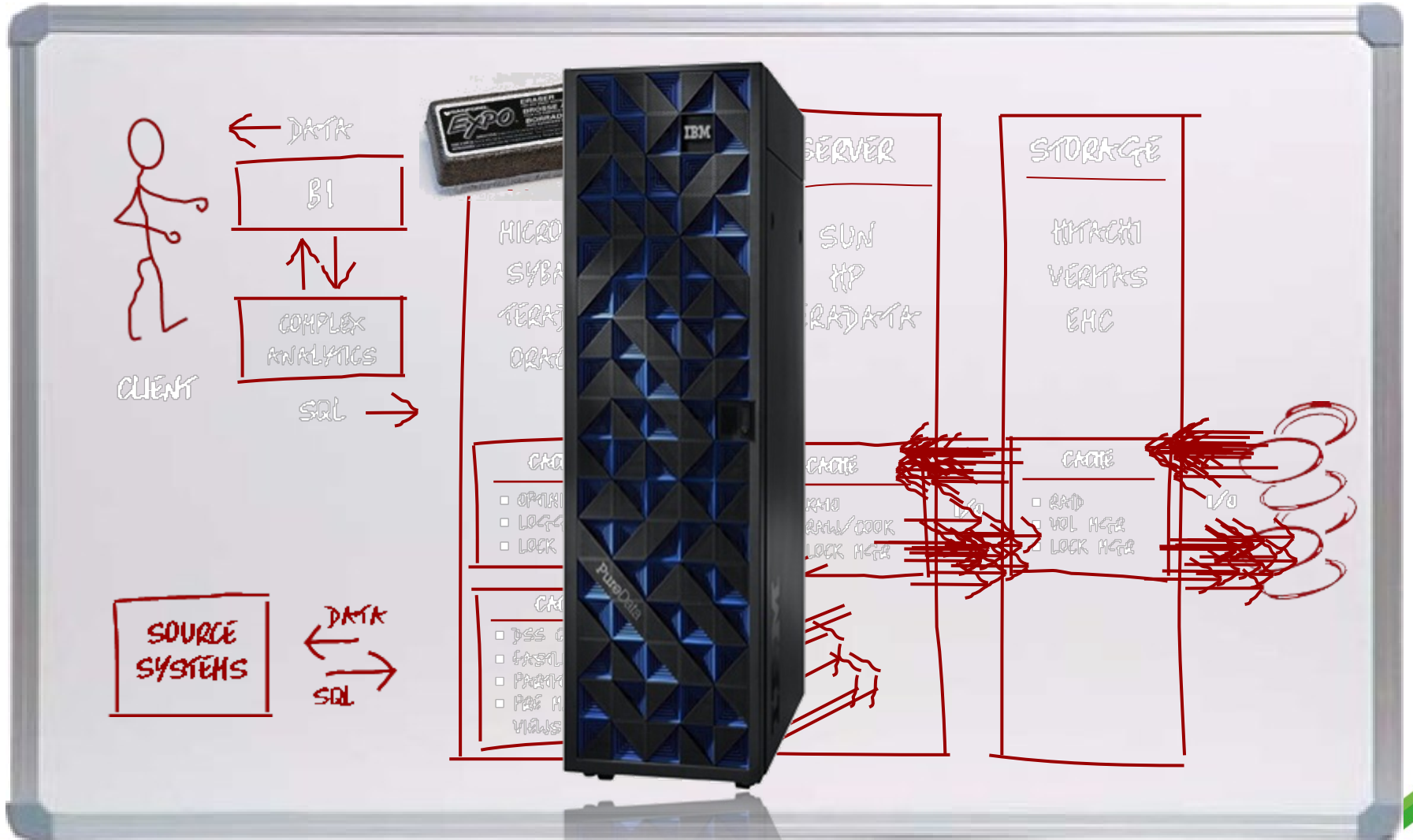
Simplified Experience

Making every part of the IT lifecycle easier - with integrated management of the entire system and a broad open ecosystem of optimized solutions

Traditional Data Warehouse Complexity



Let's Simplify This...



... Move Analytics into the Warehouse with a TRUE Appliance



The diagram on the whiteboard illustrates a data architecture. On the left, a stick figure labeled 'CLIENT' interacts with a 'BI' box. An arrow labeled 'DATA' points from the BI box to the client. Below the BI box is a 'COMPLEX ANALYTICS' box, with a double-headed arrow between them. An arrow labeled 'SQL' points from the client to the complex analytics box. At the bottom left, a box labeled 'SOURCE SYSTEMS' has a double-headed arrow labeled 'DATA' and 'SQL' connecting it to the complex analytics box. In the center is a tall server rack with a blue diamond-patterned front panel, labeled 'PureData' and 'IBM'. To the right of the rack, the text 'PureData System for Analytics' is written twice, with 'LESS IS MORE' written in large, white, hand-drawn letters in between.

Deep Analytics Appliance – Revolutionized Analytics

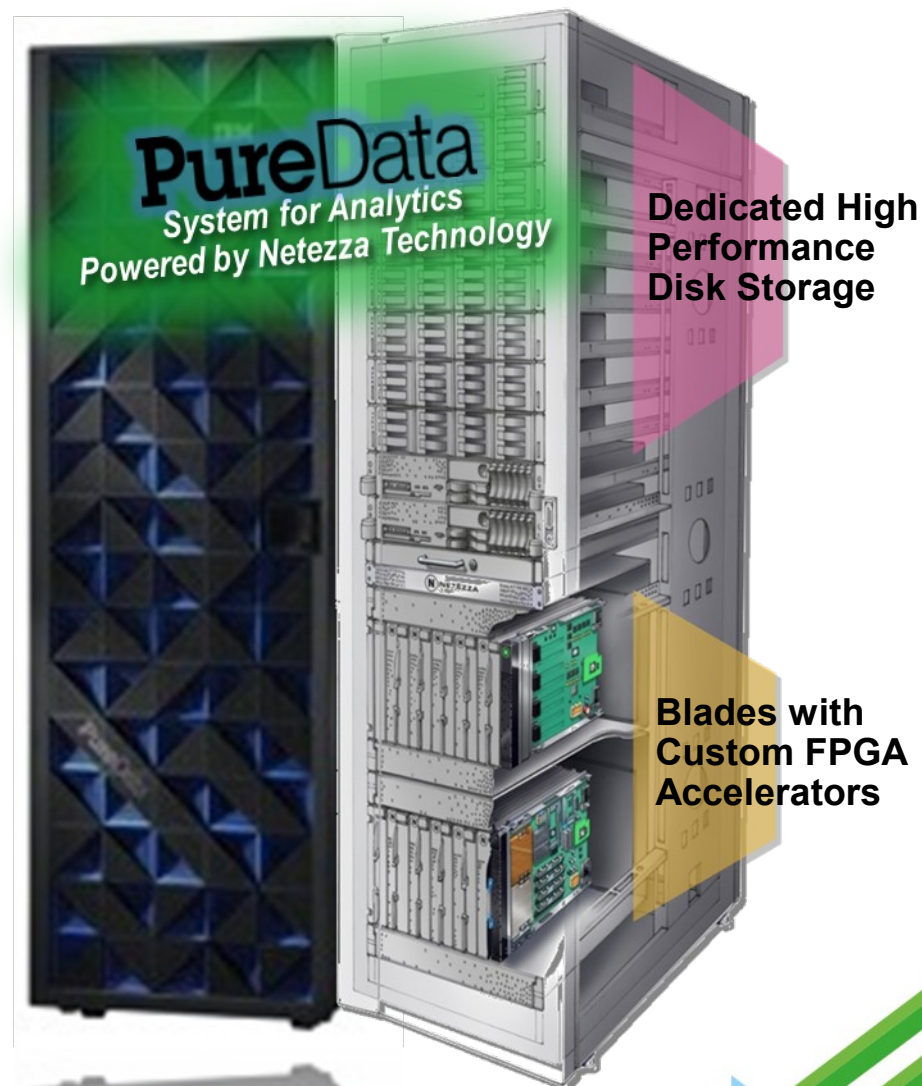
Purpose-built analytics appliance

Speed: 10-100x faster than traditional systems

Simplicity: Minimal admin/mgt. and tuning

Scalability: Peta-scale user data capacity

Smart: High-performance advanced analytics



30% increase in coupon redemption rates

70x more queries on **5x** data



Delivering personalized coupons to shoppers in real time

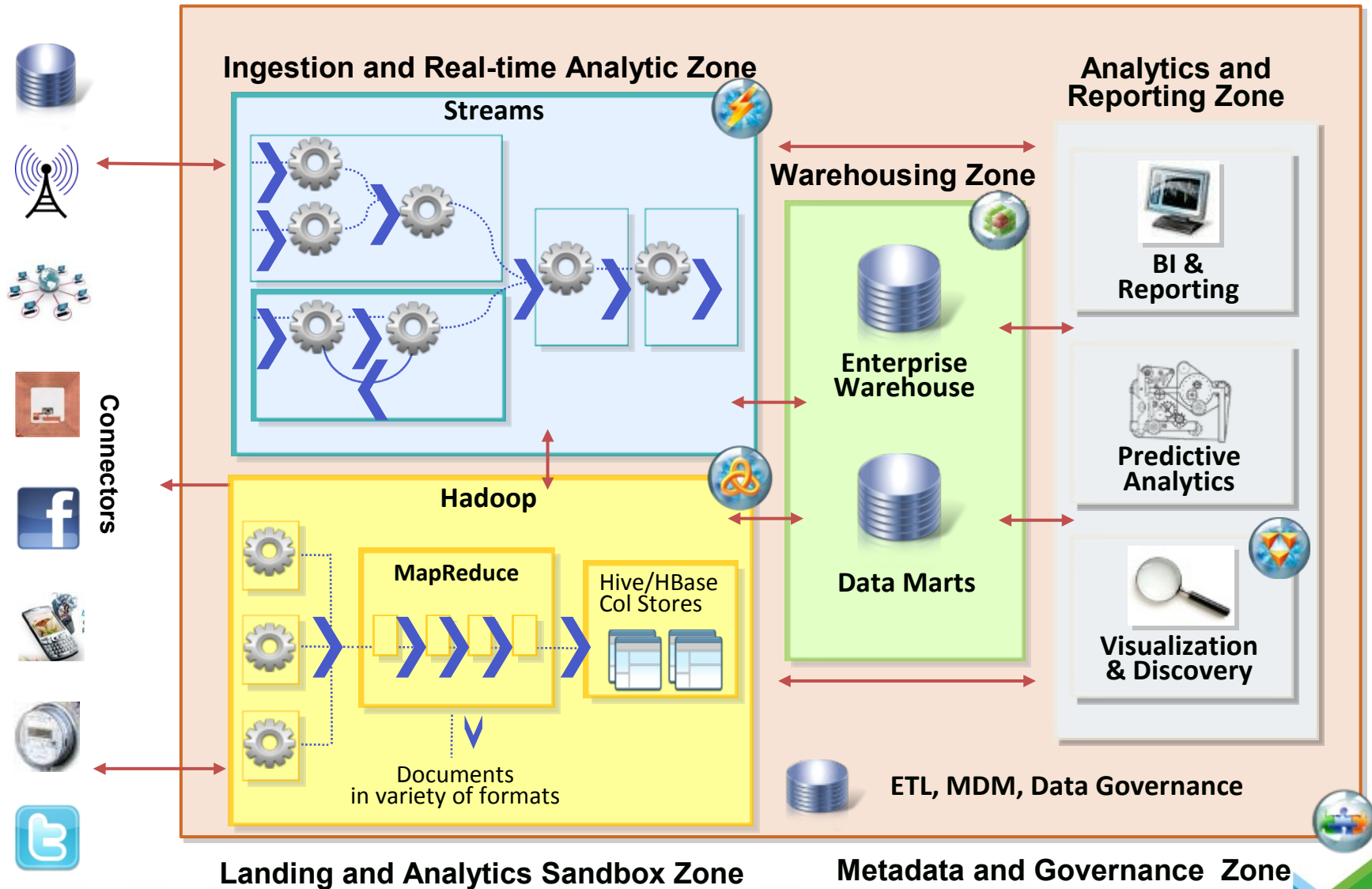
Store and access 400B market basket records to provide personalized experience

600 predictive models per year, 10X as many as before

*"Because of (Netezza's) in-database technology, we believe we'll be able to do 600 predictive models per year (**10X** as many as before) **with the same staff.**"*


CATALINA MARKETING : Williams
 and executive VP

a big data architecture



Expand from enterprise data to big data

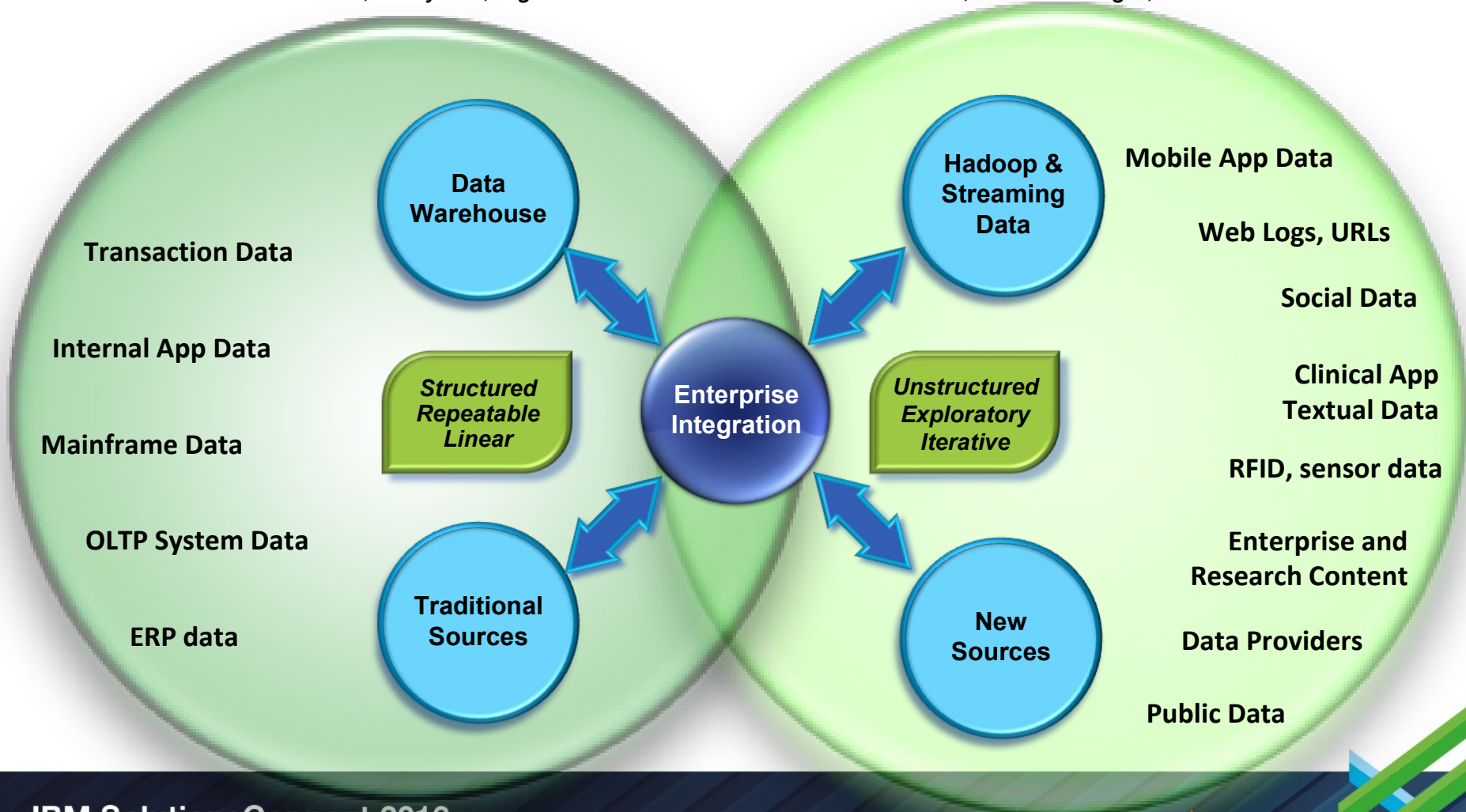


Traditional Approach

Structured, analytical, logical

New Approach

Creative, holistic thought, intuition



Start on Your Big Data Journey Today

Know more:

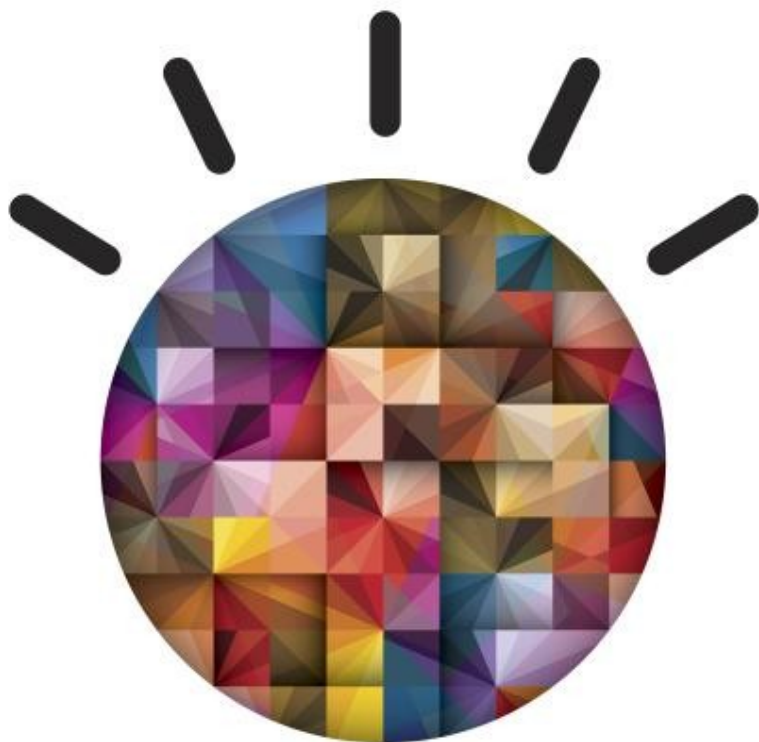
- IBM Big Data: ibm.com/bigdata
- IBMBigDataHub.com
- BigDataUniversity.com
- IBM Institute of Business Value study
- Books / analyst papers

Schedule an IBM Big Data Workshop

1. Best practices
2. Use cases
3. Business Value Assessment



For more information:
ibm.com/bigdata



- A short break...
- stay tuned for IBM's vision on how to make BigData really become the platform of trusted information!